Relatório de CT-213: Programação dinâmica

Henrique F. Feitosa

Instituto Tecnológico de Aeronáutica,
São José dos Campos, São Paulo, Brasil

1 Introdução

Nessa prática, buscou-se implementar os algoritmos de programação dinâmica no contexto da solução de um processo decisório de markov. Os algoritmos implementados foram:

1. *policy evaluation*
2. *policy iteration*
3. *value iteration*

O objetivo dessa prática é avaliar as políticas e e determinar politícas ótimas para um *grid world*.

Inicialmente, implementou-se o *policy evaluation* seguindo a equação de Bellman e usando uma condição de parada, as quais estão mostrados abaixo.

$$\nu_\pi(s) = \sum_{a \epsilon A} \pi(a|s) r(s,a) + \gamma \sum_{a \epsilon A} \sum_{s' \epsilon S} \pi(a|s) p(s'|s,a) \nu_\pi(s')$$

$$max_{s \epsilon S} |\nu_{k+1}(s) - \nu_k(s)| < \varepsilon$$

Após isso, a implementação da iteração de política alterna entre a avaliação de política e o aprimoramento de política. Ademais, a implementação do algoritmo de iteração de valor consiste em iterar diretamente sobre a função valor de acordo com a equação de otimalidade de Bellman:

$$v_*(s) = max_{a \epsilon A} (r(s,a) + \gamma \sum_{s' \epsilon S} p(s'|s,a) \nu_\pi(s'))$$

Finalmente, comparou-se os resultados dos algoritmos implementados para dois casos:

$$p_c = 1.0$$
$$\gamma = 1.0$$

$$p_c = 0.8$$
$$\gamma = 0.98$$

2 Resultados e Discussão

Os resultados obtidos estão nas figuras 1 e 2.

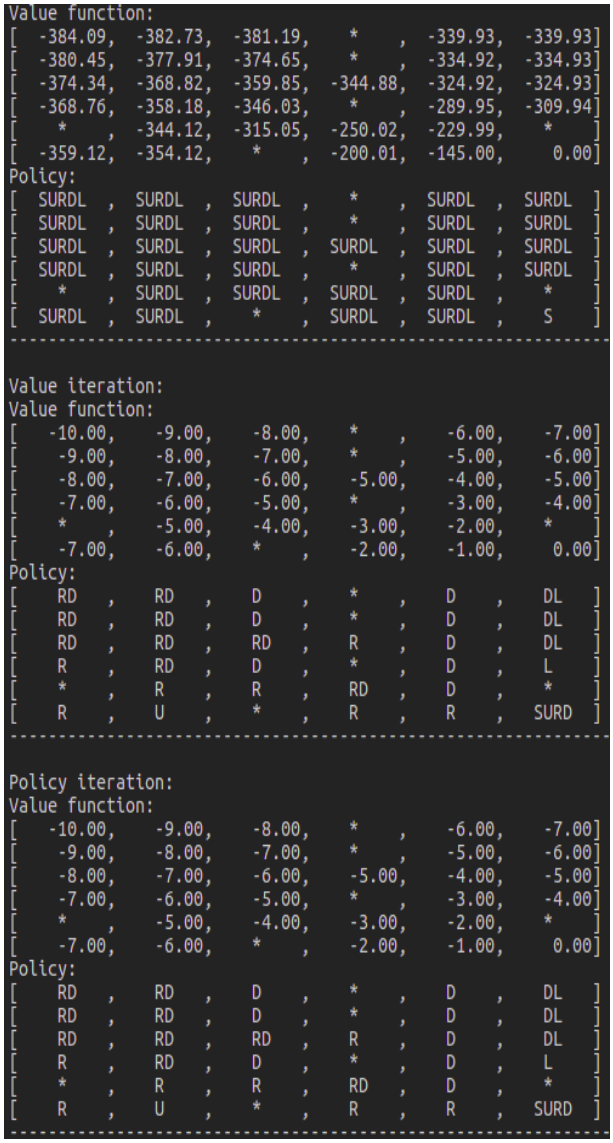

Figura 1. Mostra o resultado dos algoritmos para $\gamma = 1$ e $p_c = 1$

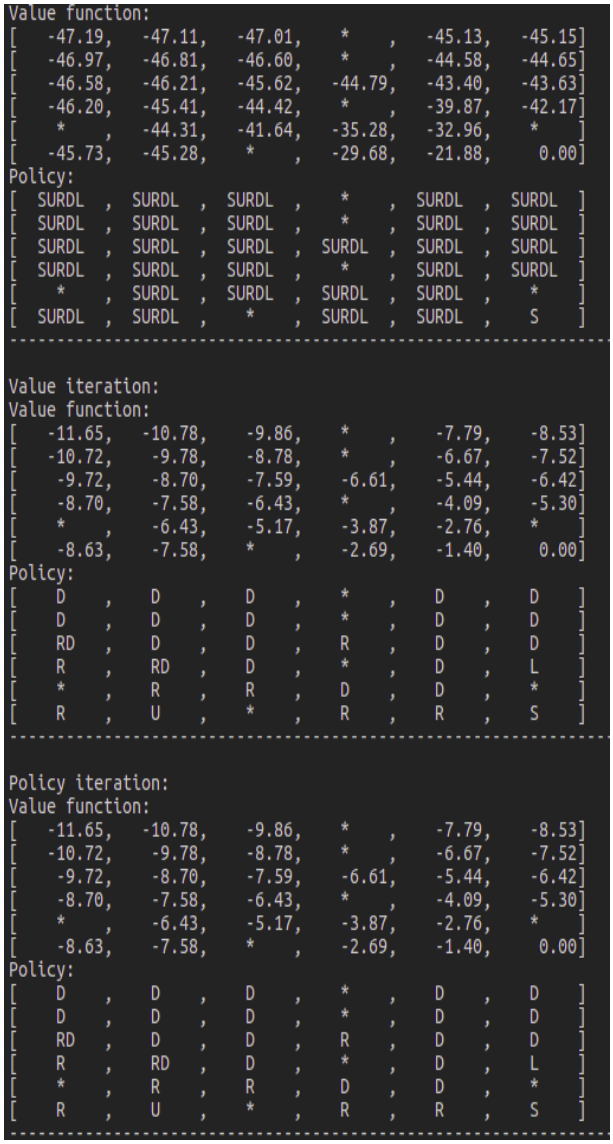

Figura 2. Mostra o resultado dos algoritmos para $\gamma = 0.98$ e $p_c = 0.8$

Assim, pode-se perceber que a política ótima encontrada foi a mesma para o *policy iteration* e para o *value iteration*, o que era espoerado. Além disso, cabe ressaltar que se diminuissemos o valor do γ e do p_c, é esperado que o *value function* diminua significantemente, uma vez que o retorno considerado se torna menor por causa do γ e o algoritmo ganha uma característica de *explotation*

com a diminuição do p_c, fazendo com que diminua a possibilidade de cair em mínimos locais.